Discriminant Analysis of Fresh Water Fish

Introduction

We are interested in whether simple morphological measurements on fish are adequate to correctly separate 159 fish caught in Finland’s lake Laengelmavesi into correct species categories.

Methods

For each of the seven species (bream, parkki, pike, perch, roach, smelt, and whitefish), the weight, length, height, and the width of each fish were measured. Three different length measurements are recorded: from the nose of the fish to the beginning of its tail, from the nose to the notch of its tail, and from the nose to the end of its tail. The height and width are recorded as percentages of the third length variable.

The discriminant analysis was performed in SAS using weight, the three length variables, height and width.

The assumption of normality was tested for weight and the first length variable using proc univariate. The assumption of equal variance was tested within the proc discrim procedure.

A cross validation was performed so that both the training classification error and the prediction error could be estimated. This procedure is also performed within proc discrim using the crossvalidate option (see notes).

Results and Conclusions

The tests for normality were significant, indicating that univariate and multivariate ANOVAs would not be appropriate for these data, and that the associated p-values would be inaccurate.

The test for equal variance was also significant and so SAS used the with-in covariance matrices and the quadratic discriminant function in the analysis.

The procedure made no errors during training, however during cross validation a total of 10 fish were misclassified. The species with the highest rate of misclassification was whitefish. All six individuals were misclassified, with three individuals misclassified as roach and three as perch. One out of 20 roach was misclassified as perch and 3 out 14 smelt were misclassified as perch. There was an overall misclassification rate of 0.1806.

While the training error was low, the high cross validated prediction error may imply that these morphological measurements would not do well in classifying an individual fish into the correct species.

Point Allocation: 40 possible

• Not in journal article format: -6
• Not stating variables used: -4
• Not distinguishing training error from cross validation error: -4
• No reporting training error: -4
• No reporting cross validation error: -4
• Poor grammar or presentation: -10
• No analysis performed: -12
Logistic Regression in Pain Therapy Outcomes

Introduction

The treatment of pain is a difficult endeavor. The perception of pain is highly individual, and hence, so is the perception of pain reduction. This study considers the analgesic effects of treatments on elderly patients with neuralgia.

Methods

There were two experimental treatments, A and B and one placebo treatment in this study. Patient characteristics of interest were sex, age, duration of pain before beginning treatment, and presence of pain after treatment. The variable age is in years and the variable duration is in months.

A logistic regression was performed in SAS using the patient characteristics to predict the presence/absence of pain after treatment.

Results and Conclusions

Only duration was non-significant at the 90% level, indicating that sex and age both play a role in patient recovery. The model was refit with duration removed. The parameter estimates, including the estimated odds ratios, are shown below.

<table>
<thead>
<tr>
<th>Analysis of Maximum Likelihood Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>Treatment A</td>
</tr>
<tr>
<td>Treatment B</td>
</tr>
<tr>
<td>Sex</td>
</tr>
<tr>
<td>Age</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Odds Ratio Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect</td>
</tr>
<tr>
<td>Treatment A vs P</td>
</tr>
<tr>
<td>Treatment B vs P</td>
</tr>
<tr>
<td>Sex F vs M</td>
</tr>
<tr>
<td>Age</td>
</tr>
</tbody>
</table>

Diagnostics

Plots of the pearson’s residuals and the deviance residuals look smooth, with one possible outlier, case number 17.
Interpretation

SAS is modeling the probability that pain = no, so a “success” here is the outcome of no pain. SAS automatically assigned the placebo treatment as the baseline, so the slopes for the two experimental treatments are interpreted as the increase in the probability of no pain with respect to the placebo.

The odds of a patient experiencing no pain after treatment $A$ are $e^{0.8772} = 2.404$ times that of a patient who received the placebo. The odds of no pain after receiving treatment $B$ are $e^{1.4246} = 4.156$ times that of the odds after receiving the placebo.

A female patient is 2.489 times more likely to be pain-free than a male of the same age who received the same treatment.

For every year increase in age the odds of being pain free after treatment decrease by 76.67%.

Treatment $B$ appears to be more effective in the treatment of neuralgia than treatment $A$, though both appear to be more effective than the placebo.

Point Allocation: 60 possible

- Not in journal article format: -9
- Not reporting model/parameter estimates: -6
- Not reporting odds ratios: -6
- No diagnostics: -6
- No or incorrect interpretation of parameters: -10
- No or incorrect interpretation of model results: -10
- No analysis performed: -22